Chapter 3
Basic Control Architecture

Check for updates

3.1 Open-Loop Control

Suppose a system benefits by tracking relatively slow oscillatory environmental fluctuations at frequency ω_e and ignoring much faster noisy environmental fluctuations at frequency ω_n. Assume that the system has an intrinsic daily oscillator at frequency $\omega_0 = 1$, with time measured in days. How can a system build a control circuit that uses its intrinsic daily oscillator to track slower environmental signals and ignore faster noisy signals?

We can begin by considering circuit designs that follow the cascade in Fig. 2.1b. That cascade is a single direct path from input to output, matching the cascade in Eq. 2.3. That path is an open loop because there is no closed-loop feedback.

Using the components in Fig. 2.1b, the internal oscillator is given by

$$P(s) = \frac{\omega_0}{s^2 + \omega_0^2},$$

and the external reference signal is given by

$$R(s) = \frac{\omega_e}{s^2 + \omega_e^2} + \frac{\omega_n}{s^2 + \omega_n^2},$$

the sum of one low- and one high-frequency sine wave. From Fig. 2.1b, the design goal seeks to create a preprocess controlling filter, $C(s)$, that combines with the intrinsic internal oscillator, $P(s)$, to transform the reference input, $R(s)$, into an output, $Y(s) \approx \omega_e/(s^2 + \omega_e^2)$, that fluctuates at ω_e and ignores ω_n.

In this case, we know exactly the intrinsic dynamics, $P(s)$. Thus, we can use the open-loop path in Fig. 2.1b to find a controller, $C(s)$, such that the transfer function $C(s)P(s)$ gives approximately the input–output relation that we seek between $R(s)$ and $Y(s)$. For example, by using the controller

Fig. 3.1 Bode plot of an
intrinsic oscillator, $P(s)$,
modulated by a controller,
$C(s)$, in an open loop
$L(s) = C(s)P(s)$. The gold
curves follow Eq. 3.3, in
which the actual frequency
of the internal oscillator is
$\tilde{\omega}_0 = 1.2$ rather than the
value $\omega_0 = 1$ that set the
design of the controller. The
underlying blue curves show
the outcome when the
internal oscillator frequency
matches the design
frequency, $\tilde{\omega}_0 = \omega_0 = 1$
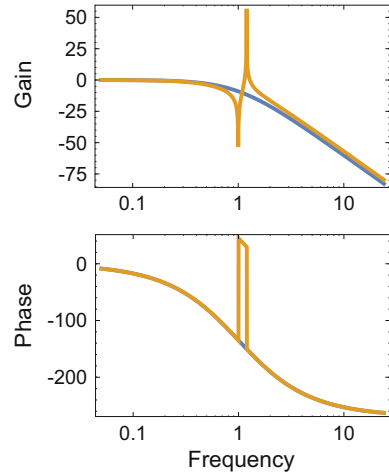
$$C(s) = \left(\frac{\omega_0}{s + \omega_0} \right)^3 \left(\frac{s^2 + \omega_0^2}{\omega_0} \right), \tag{3.1}$$

the open-loop system becomes

$$L(s) = C(s)P(s) = \left(\frac{\omega_0}{s + \omega_0} \right)^3, \tag{3.2}$$

because the second term in $C(s)$ cancels $P(s)$. The system $L(s)$ is the low-pass filter
in Eq. 2.11 raised the third power. With $\omega_0 = 1$, this system has a Bode plot similar
to the blue curve in Fig. 2.2e, f, but because of the exponent in $L(s)$, the gain falls
more quickly at high frequencies and the phase lag is greater.

As with the low-pass filter illustrated in Fig. 2.2, this open-loop system, $L(s)$,
tracks environmental signals at frequency $\omega_e \ll \omega_0$ and suppresses noisy signals at
frequency $\omega_n \gg \omega_0$. However, even if we could create this controller over the required
range of frequencies, it might turn out that this system is fragile to variations in the
parameters.

We could study robustness by using the differential equations to calculate the
dynamics for many combinations of parameters. However, such calculations are
tedious, and the analysis can be difficult to evaluate for more than a couple of param-
eters. Using Bode plots provides a much easier way to analyze system response under
various conditions.

Suppose, for example, that in the absence of inputs, the internal oscillator, $P(s)$,
actually fluctuates at the frequency $\tilde{\omega}_0 \neq \omega_0$. Then, the open-loop system becomes

$$L(s) = \frac{\tilde{\omega}_0}{\omega_0} \left(\frac{\omega_0}{s + \omega_0} \right)^3 \left(\frac{s^2 + \omega_0^2}{\omega_0} \right) \left(\frac{\tilde{\omega}_0}{s^2 + \tilde{\omega}_0^2} \right), \tag{3.3}$$

in which the first term adjusts the gain to be one at $s = 0$.

The gold curves in Fig. 3.1 show the Bode plot for this open loop, using $\omega_0 = 1$ and $\tilde{\omega}_0 = 1.2$. Note the resonant peak in the upper magnitude plot. That peak occurs when the input frequency matches the natural frequency of the intrinsic oscillator, $\tilde{\omega}_0$. Near that resonant frequency, the system "blows up," because the denominator in the last term, $s^2 + \tilde{\omega}_0^2$, goes to zero as $s = j\omega \to j\tilde{\omega}_0$ and $s^2 \to -\tilde{\omega}_0^2$.

In summary, open-loop control works well when one has accurate information. Successful open-loop control is simple and has relatively low cost. However, small variations in the intrinsic process or the modulating controller can cause poor performance or instabilities, leading to system failure.

3.2 Feedback Control

Feedback and feedforward have different properties. Feedforward action is obtained by matching two transfer functions, requiring precise knowledge of the process dynamics, while feedback attempts to make the error small by dividing it by a large quantity.
—Åström and Murray (2008, p. 320)

Feedback often solves problems of uncertainty or noise. Human-designed systems and natural biological systems frequently use feedback control.

Figure 2.1c shows a common form of negative feedback. The output, y, is returned to the input. The output is then subtracted from the environmental reference signal, r. The new system input becomes the error between the reference signal and the output, $e = r - y$.

In closed-loop feedback, the system tracks its target reference signal by reducing the error. Any perturbations or uncertainties can often be corrected by system dynamics that tend to move the error toward zero. By contrast, a feedforward open loop has no opportunity for correction. Feedforward perturbations or uncertainties lead to uncorrected errors.

In the simple negative feedback of Fig. 2.1c, the key relation between the open-loop system, $L(s) = C(s)P(s)$, and the full closed-loop system, $G(s)$, is

$$G(s) = \frac{L(s)}{1 + L(s)}. \tag{3.4}$$

This relation can be derived from Fig. 2.1c by noting that, from the error input, $E(s)$, to the output, $Y(s)$, we have $Y = LE$ and that $E = R - Y$. Substituting the second equation into the first yields $Y = L(R - Y)$. Solving for the output Y relative to the input R, which is $G = Y/R$, yields Eq. 3.4.

The error, E, in response to the environmental reference input, R, can be obtained by a similar approach, yielding

$$E(s) = \frac{1}{1 + L(s)} R(s). \qquad (3.5)$$

If the open loop, $L(s)$, has a large gain, that gain will divide the error by a large number and cause the system to track closely to the reference signal. A large gain for $L = CP$ can be achieved by multiplying the controller, C, by a large constant, k. The large gain causes the system to respond rapidly to deviations from the reference signal.

Feedback, with its powerful error correction, typically provides good performance even when the actual system process, P, or controller, C, differs from the assumed dynamics. Feedback also tends to correct for various types of disturbances and noise, and can also stabilize an unstable open-loop system.

Feedback has two potential drawbacks. First, implementing feedback may require significant costs for the sensors to detect the output and for the processes that effectively subtract the output value from the reference signal. In electronics, the implementation may be relatively simple. In biology, feedback may require various additional molecules and biochemical reactions to implement sensors and the flow of information through the system. Simple open-loop feedforward systems may be more efficient for some problems.

Second, feedback can create instabilities. For example, when $L(s) \rightarrow -1$, the denominator of the closed-loop system in Eq. 3.4 approaches zero, and the system blows up. For a sinusoidal input, if there is a frequency, ω, at which the magnitude, $|L(j\omega)|$, is one and the phase is shifted by one-half of a cycle, $\phi = \pm\pi = \pm 180°$, then $L(j\omega) = -1$.

The problem of phase arises from the time lag (or lead) between input and feedback. When the sinusoidal input is at a peak value of one, the output is shifted to a sinusoidal trough value of minus one. The difference between input and output combines in an additive, expansionary way rather than providing an error signal that can shrink toward an accurate tracking process. In general, time delays in feedback can create instabilities.

Instabilities do not require an exact half cycle phase shift. Suppose, for example, that the open loop is

$$L(s) = \frac{k}{(s + 1)^3}.$$

This system is stable, because its eigenvalues are the roots of the polynomial in the denominator, in this case $s = -1$, corresponding to a strongly stable system. The closed loop has the transfer function

$$G(s) = \frac{L(s)}{1 + L(s)} = \frac{k}{k + (s + 1)^3},$$

which has an eigenvalue with real part greater than zero for $k > 8$, causing the system to be unstable. An unstable system tends to explode in magnitude, leading to system failure or death.

3.3 Proportional, Integral, and Derivative Control

Open loop systems cannot use information about the error difference between the target reference input and the actual output. Controllers must be designed based on information about the intrinsic process and the likely inputs.

By contrast, feedback provides information about errors, and controller design focuses primarily on using the error input. Given the error, the controller outputs a new command reference input to the intrinsic system process. Precise knowledge about the intrinsic system dynamics is much less important with feedback because the feedback loop can self-correct.

This section discusses controller design for feedback systems. A controller is a process that modulates system dynamics. For the simplest feedback shown in Fig. 2.1c, we start with an intrinsic process, $P(s)$, and end up with feedback system dynamics

$$G(s) = \frac{C(s)P(s)}{1 + C(s)P(s)} = \frac{L(s)}{1 + L(s)},$$

in which $C(s)$ is the controller. The problem is how to choose a process, $C(s)$, that balances the tradeoffs between various measures of success, such as tracking the reference input and robustness to perturbations and uncertainties.

Figure 3.2a includes two kinds of perturbations. The input d describes the load disturbance, representing uncertainties about the internal process, $P(s)$, and disturbances to that internal process. Traditionally, one thinks of d as a relatively low-frequency perturbation that alters the intrinsic process. The input n describes perturbations that add noise to the sensor that measures the process output, η, to yield the final output, y. That measured output, y, is used for feedback into the system.

To analyze alternative controller designs, it is useful to consider how different controllers alter the open-loop dynamics, $L(s) = C(s)P(s)$. How does a particular change in the controller, $C(s)$, modulate the intrinsic dynamics, $P(s)$?

First, we can simply increase the gain by letting $C(s) = k_p > 1$, a method called proportional control. The system becomes $G = k_p P/(1 + k_p P)$. For large k_p and positive $P(s)$, the system transfer function is $G(s) \to 1$, which means that the system output tracks very closely to the system input. Proportional control can greatly improve tracking at all frequencies. However, best performance often requires tracking low-frequency environmental inputs and ignoring noisy high-frequency inputs from the reference signal. In addition, large k_p values can cause instabilities, and it may be that $P(s) < 0$ for some inputs.
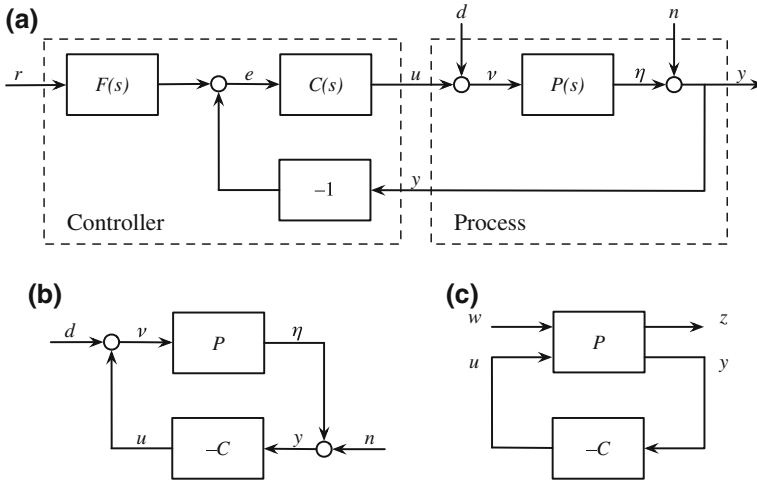
Fig. 3.2 Closed-loop feedback. **a** An extended feedback loop with inputs for disturbance, d, and noise, n. The function $F(s)$ may be used to filter the reference input, providing a second degree of freedom in addition to the main controller, $C(s)$. The system can be divided into intrinsic processes that cannot be adjusted directly and designed processes of control that can be adjusted. Note the inputs for each block: r and y for the controller, and u, d, and n for the process. **b** In this panel, the blocks P and C represent the multicomponent process and control blocks from the upper panel. The reference signal is assumed to be zero, allowing one to focus on the roles of disturbance and noise in relation to system stability. **c** An abstraction of the feedback process, in which the vector y includes all the signals from the process to the controller, u includes all the control input signals to the process, w includes all the extrinsic inputs, and z includes any additional signal outputs from the process. Redrawn from Åström and Murray (2008), © Princeton University Press

Second, we can add integral control by including the term k_i/s to the controller. We can understand why this term is an integrator by considering a few steps of analysis that extend earlier equations. Multiplying Eq. 2.5 by $1/s$ increases the denominator's order of its polynomial in s. That increase in the exponents of s corresponds to an increase in the order of differentiation for each term on the left side of Eq. 2.4, which is equivalent to integrating each term on the right side of that equation. For example, if we start with $\dot{x} = u$ and then increase the order of differentiation on the left side, $\ddot{x} = u$, this new expression corresponds to the original expression with integration of the input signal, $\dot{x} = \int u \, \mathrm{d}t$.

Integrating the input smooths out high-frequency fluctuations, acting as a filter that passes low-frequency inputs and blocks high-frequency inputs. Integration causes a slower, smoother, and often more accurate adjustment to the input signal. A term such as $a/(s + a)$ is an integrator for large s and a pass-through transfer function with value approaching one for small s.

Perfect tracking of a constant reference signal requires a pure integrator term, $1/s$. A constant signal has zero frequency, $s = 0$. To track a signal perfectly, the system transfer function's gain must be one so that the output equals the input. For the simple closed loop in Eq. 3.4, at zero frequency, $G(0)$ must be one. The tracking error is $1 - G = 1/(1 + L)$. The error goes to zero as the gain of the open loop goes to infinity, $L(0) \to \infty$. A transfer function requires a term $1/s$ to approach infinity as s goes to zero. In general, high open loop gain leads to low tracking error.

Third, we can add derivative control by including the term $k_d s$. We can understand why this term differentiates the input term by following the same steps as for the analysis of integration. Multiplying Eq. 2.5 by s increases the numerator's order of its polynomial in s. That increase in the exponents of s corresponds to an increase in the order of differentiation for each term on the right side of Eq. 2.4. Thus, the original input term, $u(t)$, becomes the derivative with respect to time, $\dot{u}(t)$.

Differentiating the input causes the system to respond to the current rate of change in the input. Thus, the system responds to a prediction of the future input, based on a linear extrapolation of the recent trend.

This leading, predictive response enhances sensitivity to short-term, high-frequency fluctuations and tends to block slow, low-frequency input signals. Thus, differentiation acts as a high-pass filter of the input signal. A term such as $s + a$ multiplies signals by a for low-frequency inputs and multiplies signals by the increasing value of $s + a$ for increasingly high-frequency inputs. Differentiators make systems very responsive, but also enhance sensitivity to noisy high-frequency perturbations and increase the tendency for instability.

A basic proportional, integral, and derivative (PID) controller has the form

$$C(s) = k_p + \frac{k_i}{s} + k_d s = \frac{k_d s^2 + k_p s + k_i}{s}. \tag{3.6}$$

PID controllers are widely used across all engineering applications. They work reasonably well for many cases, they are relatively easy to understand, and their parameters are relatively easy to tune for various tradeoffs in performance.

3.4 Sensitivities and Design Tradeoffs

Figure 3.2a shows a basic feedback loop with three inputs: the reference signal, r, the load disturbance, d, and the sensor noise, n. How do these different signals influence the error between the reference signal and the system output? In other words, how sensitive is the system to these various inputs?

To derive the sensitivities, define the error in Fig. 3.2a as $r - \eta$, the difference between the reference input, r, and the process output, η (Åström and Murray 2008, Sect. 11.1). To obtain the transfer function between each input and output, we use the rule for negative feedback: The transfer function between the input and output is the

open loop directly from the input to the output, L, divided by one plus the pathway around the feedback loop, $1 + L$.

If we assume in Fig. 3.2a that there is no feedforward filter, so that $F = 1$, and we define the main open loop as $L = CP$, then the output η in response to the three inputs is

$$\eta = \frac{L}{1 + L} r + \frac{P}{1 + L} d - \frac{L}{1 + L} n, \tag{3.7}$$

in which each term is the open loop between the input signal and the output, η, divided by one plus the pathway around the full loop, L. If we define

$$S = \frac{1}{1 + L} \qquad T = \frac{L}{1 + L} \qquad S + T = 1, \tag{3.8}$$

with S as the sensitivity function and T as the complementary sensitivity function, then the error is

$$r - \eta = Sr - PSd + Tn. \tag{3.9}$$

This expression highlights the fundamental design tradeoffs in control that arise because $S + T = 1$. If we reduce T and the sensitivity to noise, we increase S. An increase in S raises the error in relation to the reference signal, r, and the error in relation to the load disturbance, d. If we reduce S, we increase T and the sensitivity to noise, n. These sensitivity tradeoffs suggest two approaches to design.

First, the sensitivities $S(s)$ and $T(s)$ depend on the input, s. Thus, we may adjust the tradeoff at different inputs. For example, we may consider inputs, $s = j\omega$, at various frequencies, ω. Sensor noise, n, often arises as a high frequency disturbance, whereas the reference input, r, and the load disturbance, d, often follow a low frequency signal. If so, then we can adjust the sensitivity tradeoff to match the common input frequencies of the signals. In particular, at low frequency for which r and d dominate, we may choose low S values whereas, at high frequency for which n dominates, we may choose low T values.

Second, we may add an additional control process that alters the sensitivity tradeoff. For example, we may use the feedforward filter, F, in Fig. 3.2a, to modulate the reference input signal. With that filter, the transfer function from the input, r, to the error output, $r - \eta$, becomes $1 - FT$. If we know the form of T with sufficient precision, we can choose $FT \approx 1$, and thus we can remove the sensitivity of the error to the reference input.

Note that adjusting the tradeoff between S and T only requires an adjustment to the loop gain, L, which usually does not require precise knowledge about the system processes. By contrast, choosing F to cancel the reference input requires precise information about the form of T and the associated system processes. In other words, feedback is relatively easy and robust because it depends primarily on adjusting gain magnitude, whereas feedforward requires precise knowledge and is not robust to misinformation or perturbation.